Systems Approaches to Map In Vivo RNA–Protein Interactions in *Arabidopsis thaliana*

Martin Lewinski and Tino Köster

Contents

Abstract Proteins that specifically interact with mRNAs orchestrate mRNA processing steps all the way from transcription to decay. Thus, these RNA-binding proteins represent an important control mechanism to double check which proportion of nascent pre-mRNAs is ultimately available for translation into distinct proteins. Here, we discuss recent progress to obtain a systems-level understanding of in vivo RNA–protein interactions in the reference plant *Arabidopsis thaliana* using protein-centric and RNA-centric methods as well as combined protein binding site and structure probing.

M. Lewinski · T. Köster (✉)
RNA Biology and Molecular Physiology, Bielefeld University, Bielefeld, Germany
e-mail: tino.koester@uni-bielefeld.de

1 Introduction

RNA-binding proteins (RBPs) are a diverse class of proteins that control every step of RNA processing and RNA function in the cell. They are characterized by dedicated domains involved in RNA binding and can have accessory domains engaged in protein-protein interactions or enzymatic activities.

In higher plants, RBP function so far has been best studied in the reference plant *Arabidopsis thaliana*. Among the RBPs present in the *Arabidopsis* genome are 197 proteins with an RNA recognition motif (RRM), the most abundant type of RNA-binding domain, and 28 K homology (KH) domain proteins first identified in mammalian heterogeneous nuclear protein hnRNP K (Silverman et al. 2013). In addition, 26 Pumilio (PUM) domain proteins, nine DEAD-box helicases as well as five proteins with cold shock domains (CSDs) have been identified (Silverman et al. 2013). Another 450 proteins harbor pentatricopeptide repeat (PPR) domains. PPR domains consist of multiple 35-amino acid repeats of which two are known to be engaged in specific RNA recognition (Barkan and Small 2014). These proteins are imported into mitochondria or chloroplasts and regulate all aspects of RNA metabolism, e.g., RNA editing, splicing, RNA cleavage, and translation in organelles (Schmitz-Linneweber and Small 2008; Barkan and Small 2014).

A suite of Arabidopsis RBPs have been experimentally characterized, mainly through loss-of-function mutants and transgenic plants ectopically overexpressing RBPs. These approaches revealed a crucial role for RBPs in development (Kalyna et al. 2003; Ripoll et al. 2006; Kupsch et al. 2012; Völz et al. 2012; Ferrari et al. 2017; Foley et al. 2017; Teubner et al. 2017), timing of plant reproduction (Macknight et al. 1997; Streitner et al. 2008; Hornyik et al. 2010), responses to abiotic stress (Kim et al. 2007b, c, 2008, 2010; Park et al. 2009), pathogen defense (Fu et al. 2007; Qi et al. 2010; Jeong et al. 2011; Lyons et al. 2013; Nicaise et al. 2013), responses to phytohormones (Lu and Fedoroff 2000; Hugouvieux et al. 2001; Riera et al. 2006; Carvalho et al. 2010; Hackmann et al. 2014; Löhr et al. 2014), and circadian timekeeping (Heintzen et al. 1994; Staiger 2001; Jones et al. 2012; Schmal et al. 2013; Perez-Santángelo et al. 2014). At the biochemical level, an impact of defined RBPs on RNA processing including pre-mRNA splicing, 3′ end processing, processing of microRNA precursors, and translation has been described (Lopato et al. 1999; Simpson et al. 2003; Vazquez et al. 2004; Dong et al. 2008; Stauffer et al. 2010; Ren et al. 2012; Rühl et al. 2012; Juntawong et al. 2013; Sorenson and Bailey-Serres 2014; Staiger 2015; Carvalho et al. 2016). Recent attempts to comprehensively identify RBPs, summarized in Sect. 2, provided experimental evidence for RNA binding for most of the previously identified Arabidopsis RBPs and identified a plethora of proteins with noncanonical RBDs.

Systems approaches to describe RNA–protein interactions globally come in two main flavors (Fig. 1). In RNA-centric approaches, proteins associated with mRNAs are recovered by RNA pull-down and identified by mass spectrometry, a technique referred to as mRNA interactome capture (Baltz et al. 2012; Castello et al. 2012) (Fig. 1a). In protein-centric approaches, the focus is laid on a particular RBP. The

RNA complement associated with the RBP of interest, the ribonome, is identified via immunoprecipitation of the RBP from cell lysates and identification of the bound target RNAs, initially by microarrays (Tenenbaum et al. 2000; Galgano and Gerber 2011; Guerreiro et al. 2014) or more recently via high throughput sequencing (Licatalosi et al. 2008; König et al. 2010; Rossbach et al. 2014; Müller-McNicoll et al. 2016) (Fig. 1b).

2   The Arabidopsis RBPome

Of all predicted RBPs in Arabidopsis, RNA binding has only been experimentally confirmed for a limited number of them. A first attempt to globally identify proteins based on their ability to interact with mRNAs in vivo was made for cultured Arabidopsis cells (Schmidt et al. 2010). In this study, mRNAs and interactors were recovered under native conditions by affinity chromatography on an oligo(dT) cellulose column followed by two-dimensional gel electrophoresis. The protein components were identified via Maldi-TOF. In the RNA-bound proteome were a suite of RRM proteins including members of the family of glycine-rich RNA-binding proteins like *At*GRP2 (*Arabidopsis thaliana* glycine rich RNA-binding protein 2), *At*GRP7 and *At*GRP8 (Lewinski et al. 2016), the two oligouridylate-specific RBP45 and RBP47 proteins (Lorkovic et al. 2000), and CSD proteins.

In 2012, mRNA interactome capture was reported to comprehensively identify proteins interacting with mRNAs in mammalian cells (Baltz et al. 2012; Castello et al. 2012). This technique employs in vivo cross-linking of mRNA and bound proteins by UV light irradiation. The RNA–protein complexes are recovered by pull-down of polyadenylated RNAs using magnetic beads coated with oligo(dT). Proteins are released by RNase treatment, subjected to tryptic digest and identified via mass spectrometry (Fig. 1a). Following these pioneering studies, this technique was applied to a wide range of organisms including yeast, *Drosophila melanogaster*, *Caenorhabditis elegans*, Leishmania, trypanosomes, and Plasmodium (Mitchell et al. 2013; Beckmann et al. 2015; Matia-Gonzalez et al. 2015; Bunnik et al. 2016; Lueong et al. 2016; Sysoev et al. 2016; Wessels et al. 2016; Nandan et al. 2017). A minimal core mRNA bound proteome occurring in both human and yeast was defined by Beckmann and coworkers (Beckmann et al. 2015). Lately, mRNA interactome capture has also been successfully applied to Arabidopsis (Marondedze et al. 2016; Reichel et al. 2016; Zhang et al. 2016).

2.1   *The mRNA Interactome of Arabidopsis Protoplasts*

The first mRNA interactome capture experiments in Arabidopsis employed widely differing tissues to catalog RBPs. Gueten and coworkers chose protoplasts, cells without a cell wall, assuming that UV cross-linking should occur as efficiently as
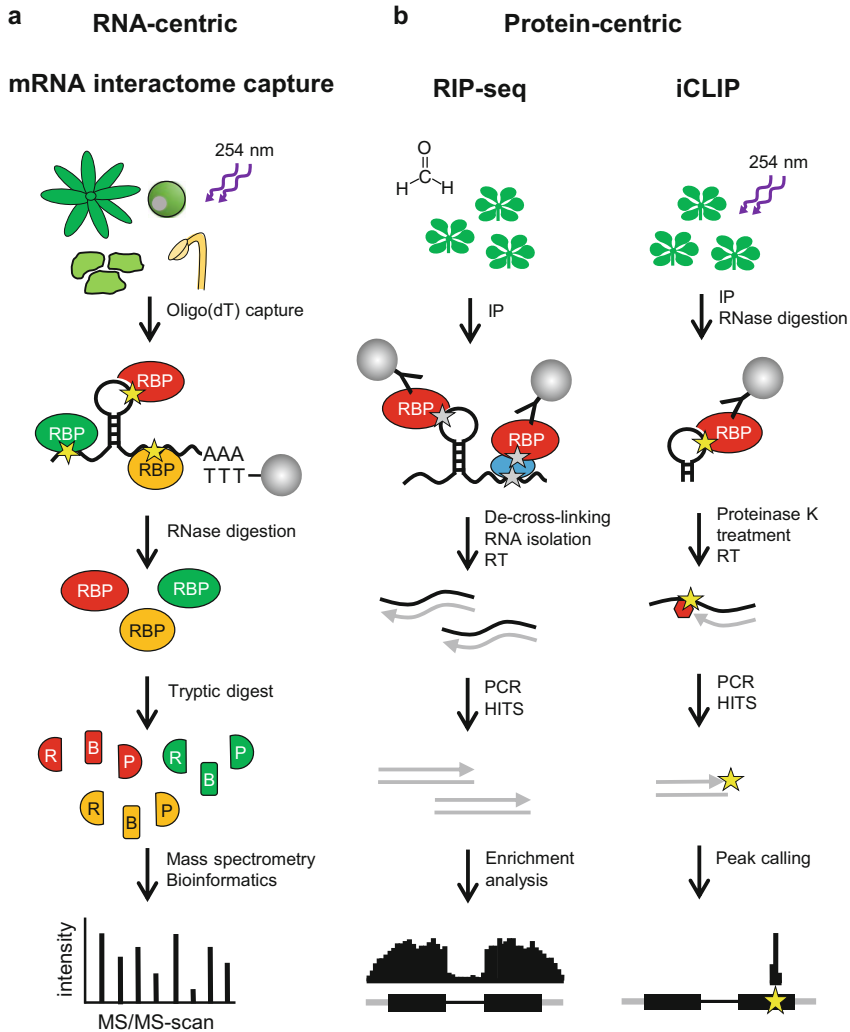
Fig. 1 Strategies to globally identify in vivo RNA–protein interaction in Arabidopsis. (**a**) RNA-centric strategies such as mRNA interactome capture employ oligo(dT) affinity capture. RNA and bound proteins are covalently linked in planta through UV irradiation. RNA–protein complexes are recovered by oligo(dT) pull-down. Proteins are released by RNase treatment, subjected to tryptic digest and identified via mass spectrometry. (**b**) Protein-centric methods focus on a particular RBP and aim at identifying its in vivo RNA targets. Based on the cross-linking agent, RIP using formaldehyde or CLIP technique using UV light are distinguished. Proteins are immunoprecipitated. In RIP-seq cross-links are reversed by heat treatment, RNA is isolated, subjected to reverse transcription and PCR amplification for HITS. Targets enriched upon RIP are determined relative to mock IP controls, e.g., Xing et al. 2015, or relative to polyadenylated RNA, e.g., Meyer et al. 2017. In iCLIP (König et al. 2010), RNA–protein complexes are subjected to RNase treatment. Bound proteins are digested with proteinase, leaving a polypeptide at the cross-link site. Reverse transcriptase stops there, allowing the detection of the cross-link site at the −1 position of the processed sequencing reads

in mammalian cell monolayers (Zhang et al. 2016). Leaf mesophyll protoplasts are also widely used in transient assays to study the regulation of gene expression.

A mesophyll protoplast mRNA interactome was defined with a total of 325 proteins based on enrichment in cross-linked samples vs. non-cross-linked controls with a $\log_2$ fold change above 2 (Zhang et al. 2016). Of these, one class was represented by 123 ribosomal proteins of which 52 were also present in the core mRNA-bound proteome of human and yeast cells (Beckmann et al. 2015). The second class comprised 70 proteins with a known RBD. For 41 of them, a role in mRNA binding and RNA biology had already been described while the remaining proteins had a potential role in mRNA processing. Moreover, 12 of the RBPs in the second class overlapped with the RBPs identified in the native oligo(dT) affinity chromatography approach (Schmidt et al. 2010). The third class comprised 132 candidate RBPs. Of these, 49 were metabolic enzymes, mainly oxidoreductases. Moreover, numerous proteins related to photosynthesis were found. As these are generally strongly expressed, their RNA binding activity and the domains involved beg for an independent validation. One of the enzymes was the Arabidopsis ortholog of phosphoglycerate kinase whose RNA binding capacity has previously been validated in yeast and human cells (Beckmann et al. 2015).

2.2   The mRNA Interactome of Etiolated Arabidopsis Seedlings

Another mRNA interactome capture experiment employed 4-days-old etiolated Arabidopsis seedlings (Reichel et al. 2016). This was based on the rationale that UV-absorbing pigments present in green plant tissue may interfere with UV cross-linking *in planta* and their absence in etiolated tissue may allow more efficient UV cross-linking.

Around 300 of the 746 proteins identified altogether were significantly enriched in UV cross-linked samples vs. non-cross-linked controls with a false discovery rate below 1% and designated the "At-RBP set." Eighty percent of these have a known RBD, and 75% have been linked to RNA biology. More than 400 additional proteins did not meet the significance criteria applied for the "At-RBP set" and were classified as "candidate RBPs."

Notably, of the 197 computationally predicted RRM proteins in Arabidopsis 160 were detected in the input fraction in etiolated seedlings (Silverman et al. 2013). Half of these were recovered in the "At-RBD set" and another 50 were present among the "candidate RBPs." Similarly, seven of the predicted KH proteins were present in the "At-RBD set" and 12 were among the "candidate RBPs." Of the predicted 450 members of the PPR protein family only 60 were detected in the input fraction, likely due to low abundance (Schmitz-Linneweber and Small 2008; Reichel et al. 2016). Only six PPR proteins were found in the "At-RBP set" and another twelve in the "candidate RBPs," likely because most RNAs in the organelles lack poly(A) tails. A comparison of the identified proteins to the mRNA interactome in other model organisms revealed that 52 were present in the interactomes of humans

(Baltz et al. 2012; Beckmann et al. 2015), mice (Kwon et al. 2013; Liao et al. 2016), and yeast (Beckmann et al. 2015) and were assigned to basic functions in RNA metabolism such as translation, splicing, and RNA unwinding.

In addition to RBPs with known RBDs many Arabidopsis proteins emerged that have not been linked to RNA binding so far. Among novel RBPs were proteins harboring a YT521-B homology (YTH) domain (Li et al. 2014). YTH domain proteins have been shown to bind N^6 methyladenosine and thus serve as readers of the m^6A mark in mammals (Wang et al. 2014). In addition, Alba domain containing proteins have been identified. Alba domain proteins are well characterized in archaebacteria where they act as transcriptional repressors and in other eukaryotes where they control translation (Goyal et al. 2016). In plants, they have not yet been functionally characterized. The only observation pointing to RNA binding is the recovery of an Arabidopsis Alba domain protein by RNA-affinity chromatography (Gosai et al. 2015). WHIRLY domain containing proteins have been characterized as single-stranded DNA binding proteins in organelles (Krause et al. 2009) and in maize, association of a WHIRLY protein with chloroplast transcripts has been observed (Prikryl et al. 2008). The identification of three WHIRLY proteins in the etiolated seedling interactome (Reichel et al. 2016) and of WHIRLY1 upon oligo(dT) affinity chromatography in Arabidopsis cells (Schmidt et al. 2010) now provides evidence for global in vivo RNA binding.

In addition, a plethora of proteins with potential RNA binding activity have been detected. To substantiate their RNA-binding properties, independent replication is desirable. Among those are proteins with the Domain of unknown function 1296, cytoskeletal proteins, and photoreceptors. The identification of plasma membrane intrinsic proteins has led to the speculation that aquaporins may be involved in transport of RNAs between cells (Reichel et al. 2016).

2.3 The mRNA Interactome of Arabidopsis Cultured Cells and Leaves of Adult Plants

Another mRNA interactome capture experiment was performed on cell suspension cultures generated from roots of the Arabidopsis accessions Col-0 and Landsberg *erecta*. In parallel, leaves of four-weeks-old Arabidopsis Col-0 plants were investigated (Marondedze et al. 2016). Of 1145 proteins identified altogether in these three samples, 914 appeared only in UV cross-linked samples, and 233 proteins were significantly enriched upon UV cross-linking relative to non-cross-linked samples. More than 350 proteins were known RBDs whereas 736 were novel candidate RBPs not previously assigned an RNA-related function or known RBD, including many enzymes of intermediary metabolism, and thus await further experimental proof (Marondedze et al. 2016).

The discovery of many novel RBPs begs for further investigation of the RNA-binding properties of these proteins. Accordingly, methods to define RNA targets of candidate RBPs genome wide using protein-centric methods have recently been adapted for the use in Arabidopsis, as discussed below.

3   Toward Arabidopsis Ribonomes

Approaches to globally identify in vivo targets of an RBP in Arabidopsis mostly rely on transgenic plants expressing an epitope-tagged version of the RBP. Immunop-urification is performed via an antibody directed against the epitope tag. To mirror-image the endogenous expression pattern, authentic promoters are used and the constructs are introduced into a loss-of-function mutant (Köster and Staiger 2014). Alternatively, endogenous RBPs can be recovered with dedicated antibodies.

To freeze the in vivo RNA–protein interactions before cell lysis, cross-linking is performed by exposing plants to formaldehyde in RNA immunoprecipitation (RIP) or by UV irradiation in UV cross-linking and immunoprecipitation (CLIP) (Fig. 1b). Formaldehyde efficiently cross-links nucleic acids and proteins in vivo but also cross-links proteins. Thus, not only direct targets are recovered. This is circumvented by using 254 nm UV light that cross-links proteins directly binding to nucleic acids in the neighborhood of the excited nucleobase but does not cross-link proteins.

To date, a comprehensive determination of in vivo targets, the ribonome, has been performed for only a few Arabidopsis RBPs, both nucleocytoplasmic proteins and chloroplast-localized proteins with different tasks in posttranscriptional regulation. In the subsequent sections, selected examples are presented.

3.1   *HLP1, An hnRNP A/B-Like Protein Involved in Alternative Polyadenylation*

HLP1 is an Arabidopsis RBP resembling mammalian hnRNP A/B-like proteins (Zhang et al. 2015). High throughput sequencing (HITS)-CLIP of HLP1 fused to GFP and expressed under control of the strong, constitutive Cauliflower Mosaic Virus 35S RNA promoter identified above 5500 transcripts bound in vivo (Zhang et al. 2015). When endogenous HLP1 protein was precipitated by a specific antibody, 6850 transcripts bound in vivo were detected with an overlap of above 3000 transcripts to the HLP1-GFP precipitation. The prevalence of cross-linked regions near polyadenylation sites provoked the hypothesis that HLP1 may control polyadenylation. Indeed, in more than 2000 transcripts the distal polyadenylation site was preferred over the proximal polyadenylation site in *hlp1* mutant plants. Around 19% of these transcripts were also recovered by HLP1 HITS-CLIP, pointing

to a role for HLP1 in the control of alternative polyadenylation, at least partly by direct binding. In line with this, MEME motifs overrepresented in the cross-link regions, namely A-rich (5′-AGAAAA-3′) and U-rich (5′-UUUUCU-3′) motifs, resembled motifs enriched in the vicinity of the poly(A) site, 5′-AAAGAAAA-3′ and 5′-UGUUUC-3′. The presence of cross-link regions in other parts of the transcripts apart from the 3′ untranslated region (UTR) suggests that HLP1 may also affect other aspects of pre-mRNA processing in addition to polyadenylation.

3.2   The Glycine-Rich RBP AtGRP7

*At*GRP7 (*Arabidopsis thaliana* glycine rich RNA-binding protein 7) is another hnRNP-like protein with an N-terminal RRM and a C-terminus enriched in contiguous glycine residues. *At*GRP7 is regulated by the circadian clock and negatively autoregulates its own oscillations by alternative splicing and Nonsense-mediated decay (Staiger et al. 2003; Schmal et al. 2013). Additionally, it is involved in several steps of posttranscriptional regulation including alternative splicing, nucleic acid chaperone function, and pri-miRNA processing (Kim et al. 2007a; Streitner et al. 2012; Köster et al. 2014). To gain insights into the breadth of its in vivo targets, individual nucleotide resolution cross-linking and immunoprecipitation (iCLIP) and RIP-seq were performed (Meyer et al. 2017). *At*GRP7 fused to GFP was expressed from its own promoter including all regulatory elements (5′ UTR, intron, and 3′ UTR) in the *atgrp7-1* loss-of-function mutant. In parallel, transgenic plants expressing GFP alone or an RNA-binding dead variant of *At*GRP7 with a single conserved arginine in the RRM mutated to glutamine (*At*GRP7 R^{49}Q) were used as negative controls.

iCLIP identified 858 transcripts with significant iCLIP hits in four out of five biological replicates for *At*GRP7-GFP that were not present in the controls. RIP-seq identified 2453 transcripts enriched by *At*GRP7-GFP relative to total polyadenylated RNA. The higher number may be due to the higher cross-linking efficiency of formaldehyde compared to UV light, and the recovery of many indirect targets. 452 transcripts were common in both data sets, suggesting that they represent a set of high confidence binders. The iCLIP cross-link sites were observed in all transcript regions, the UTRs, coding sequence and introns. After correcting for the length of the feature in the genome, cross-link sites in the 3′ UTR prevailed. Conserved motifs in the vicinity of the cross-link sites generally were U/C rich.

To determine how *AtGRP7* may impact its downstream targets, the binding targets were cross-referenced against transcriptome data from *At*GRP7 overexpressing plants or loss-of-function mutants. In both, the *At*GRP7 overexpressors or the mutant, a similar number of transcripts was expressed at elevated or reduced levels compared to wild-type plants. Notably, significantly more differentially expressed iCLIP targets were downregulated in *At*GRP7-overexpressors than upregulated. In turn, more of the differentially expressed *At*GRP7 iCLIP targets were expressed at elevated in the mutant than at reduced levels. This indicates a predominantly

negative effect of *At*GRP7 on its targets. Among the targets were more circadianly regulated transcripts than expected. In particular, elevated *At*GRP7 levels lead to damping of circadian oscillations of target transcripts including *DORMANCY/AUXIN ASSOCIATED FAMILY PROTEIN2* and *CCR-LIKE*. This conforms with the idea that the circadian clock regulated *At*GRP7 functions as a molecular slave oscillator, conveying temporal information from the core circadian clock within the cell (Rudolf et al. 2004). In addition, changes in splicing patterns were observed for iCLIP and RIP-seq targets upon misexpression of *At*GRP7, confirming a role for *At*GRP7 in the control of alternative splicing.

3.3   The Splicing Regulator SR45

Arabidopsis thaliana serine/arginine rich (SR)-like protein SR45, the counterpart of metazoan RNPS1, is an SR-like protein with two RS domains, flanking either side of the RRM (Badolato et al. 1995; Golovkin and Reddy 1999). Notably, recombinant Arabidopsis SR45 can activate splicing of a β-globin splicing reporter in HeLa cell S100 extracts (Ali et al. 2007). SR45 occurs in two splice isoforms that arise through differential usage of a 3′ splice site in intron 6. This leads to two protein isoforms that differ by seven amino acid residues and in their function: SR45.1 is involved in petal development in flowers, whereas SR45.2 is important for root growth (Zhang and Mount 2009). Genome-wide targets for SR45.1 were determined during early seedling development (Xing et al. 2015) and in inflorescences (Zhang et al. 2017), respectively.

In seedlings, RIP-seq identified 4361 transcripts from 4262 genes that were enriched upon precipitation of SR45.1-GFP from nuclei of transgenic plants compared to mock precipitation from wild type plants (Xing et al. 2015). These were designated SARs, for SR45 associated RNAs. A Gene Ontology term analysis showed that 43 of 147 abscisic acid (ABA) signaling genes (30%) were among the SARs, in line with a function for SR45 in the ABA signaling pathway (Carvalho et al. 2010). Hundred and forty-eight of the SARs had an altered expression in the *sr45-1* mutant, suggesting that binding of SR45 has functional consequences.

A MEME search for SR45 binding motifs revealed four overrepresented motifs within SAR genes. Two G/A rich motifs are largely positioned within exons and show strong similarity to the binding motifs of two metazoan splicing regulators Transformer 2 (Tra2) and serine/arginine-rich splicing factor 10 (SRSF10). Furthermore, one G/A rich motif closely resembles the GAAG motif, a known *cis*-regulatory element in regulating alternative splicing in plants. In contrast, two U/C rich motifs peak within intronic regions near 5′ and 3′ splice sites, in line with the observation that the majority of SARs were from intron-containing genes and the known role as a splicing regulator (Xing et al. 2015).

To gain insights into a potential role of SR45 in flower development, RIP-seq was performed for SR45.1-GFP in inflorescence tissue (Zhang et al. 2017). The resulting reads were analyzed by two different bioinformatics pipelines, one based

on mapping reads to the genome and one directly quantifying annotated transcripts. SARs in inflorescence were defined based on a twofold enrichment compared to GFP only controls and the identification by both pipelines. Of 1812 SARs in inflorescence, 677 overlapped with the SARs in seedlings.

Notably, 19 transcripts encoding splicing factors were among the SARs including SR45 itself, the three SR proteins SR30, SR34, and SCL35, the pre-mRNA processing factors PRP39, PRP40A, PRP40B, and PRP2, and the RNA helicase RH42, pointing to a hierarchical regulation of posttranscriptional regulators (Keene 2007). Genes upregulated in the *sr45-1* mutant are enriched for defense response genes. Indeed, the *sr45-1* mutant was more resistant to bacterial and fungal pathogens. Of 68 upregulated defense response genes in *sr45-1*, 10 were SARs. Thus, SR45 has an additional role as a negative regulator of plant immunity.

Furthermore, 81 of the inflorescence SARs were aberrantly spliced in the *sr45-1* mutant. Determination of potential SR45 binding sites in inflorescence SARs uncovered an overrepresentation of the purine-rich motifs GGNGG, GNGGA, and GNGGNNG. Importantly, GGNGG and related motifs are enriched in introns and exons that are alternatively spliced in the *sr45-1* mutant, irrespective of the splicing event is favored or suppressed by SR45. This led to the suggestion that SR45 identifies regions for alternative splicing and acts as a facilitator for other splicing factors. However, the identified binding motifs for SR45 in inflorescences differ from that in seedlings, which might be in part due to the different bioinformatic tools used for motif determination. Both RIP-seq data sets nevertheless strengthen SR45's key role as an important splicing factor in Arabidopsis. However, in both RIP-seq experiments intron-less transcripts were identified in addition to intron-containing transcripts, pointing to functions of SR45 beyond its known role in pre-mRNA splicing.

Interestingly, a comparison between the U/C-rich motifs of *At*GRP7 and the U/C-rich motifs of SR45 identified by MEME in seedlings revealed a high degree of similarity (Meyer et al. 2017). The functional significance remains to be tested.

3.4   Cold Shock Protein 1

In bacteria, CSPs are upregulated upon cold stress and destabilize RNA secondary structure at low temperatures (Sommerville 1999). To elucidate a potential involvement of Arabidopsis CSPs in the regulation of cold responsive genes, RIP followed by gene chip analysis was performed for CSP1 (Juntawong et al. 2013).

More than 6000 mRNAs were identified. Comparison of these CSP1-associated transcripts in total RNA and RNA loaded onto polysomes revealed an enrichment of mRNAs associated with ribosome biogenesis in the pool of actively translating RNAs. The high GC content in 5′ UTRs of these mRNAs suggested that CSP1 is involved in removing secondary structures in the 5′ UTR to facilitate their translation. Accordingly, these mRNAs were less efficiently loaded onto polysomes

at low temperature in the *atcsp1-1* mutant compared to wild type plants or CSP1 overexpressing plants (Juntawong et al. 2013).

3.5 The cpRNP Family

The highly abundant chloroplast ribonucleoproteins (cpRNPs) have been well characterized for their role in regulating chloroplast transcripts (Ohta et al. 1995). The cpRNPs comprise an acidic domain and two RRMs. They are encoded in the nucleus and imported into chloroplasts. Mutants in distinct cpRNPs are widely affected in processing of transcripts in the chloroplast, leading to defects in chloroplast development and, consequently, plant performance owing to the essential role of the chloroplast in photosynthetic energy (Ruwe et al. 2011). For example, mutants deficient in CP29A (29 kDa chloroplast protein A) and CP31A (31 kDa chloroplast protein A) showed gross defects at low ambient temperature. RIP performed with antibodies against the endogenous proteins and subsequent hybridization of coprecipitated RNAs on tiling arrays covering the Arabidopsis chloroplast genome (RIP-Chip) showed that CP29A and CP31A associate with large overlapping sets of chloroplast transcripts including strong enrichment for *psbB, psbD, psaA/B, atpB, ndhB* and intermediate enrichment for almost all chloroplast mRNAs (Kupsch et al. 2012). Both CP29A and CP31A are required for accumulation of chloroplast mRNAs under cold stress. Furthermore, binding of CP31A to 3′ ends of certain transcripts serves to protect these transcripts against 3′ exonuclease activity (Kupsch et al. 2012). Together with the known role of CP31A in RNA (Tillich et al. 2009) this points to multiple functions in posttranscriptional regulation in chloroplasts.

For CP33A (33 kDa chloroplast protein A), RIP-chip revealed an association with a large body of chloroplast mRNAs (Teubner et al. 2017). A global reduction in mRNAs and proteins making up the photosynthetic apparatus was found in the *cp33a* mutant. In line with a crucial role for CP33A in the development of the photosynthetic apparatus, *cp33a* null mutants have an albino phenotype and are not able to survive without external sucrose supply (Teubner et al. 2017).

3.6 The PPR Protein AtCPR1

In contrast to the broad substrate specificity of the cpRNPs, a very narrow substrate specificity was found for a representative of the PPR class of nuclear-encoded RBPs that are imported into organelles. *At*CPR1 (*Arabidopsis thaliana* CHLOROPLAST RNA PROCESSING 1) is important for the production of subunits of the thylakoid protein complexes (Ferrari et al. 2017). *Atcpr1* mutants are yellow-white because the subunits of the photosynthetic apparatus do not accumulate.

RIP-chip was performed for *At*CPR1 under native conditions. Hybridization of bound targets to chloroplast tiling arrays revealed specific binding of *At*CPR1 to only few transcripts, the *psaC* transcript encoding a photosystem I subunit, *petB-petD* encoding Cytochrome b_6 and the subunit IV of the cytochrome b_6/f complex. Because during RIP RNase was used to digest unprotected RNA, it was possible to delineate the binding regions. Binding to the *petB-petD* intergenic region correlated with a requirement for processing of the polycistronic transcript comprising *petB* and *petD* (Ferrari et al. 2017), thus providing proof for the functional relevance of the observed in vivo binding.

4 Combined Analysis of RNA–Protein Interaction and RNA Secondary Structure Landscapes

In addition to RNA sequence, RNA secondary structure also strongly influences the interaction of RBPs with their cognate RNA binding motifs (Cruz and Westhof 2009; Vandivier et al. 2016). RNA structure may facilitate binding of RBDs with a preference for double-stranded RNA or inhibit binding of RBPs with a preference for single-stranded RNA. Protein interaction profile sequencing (PIP-seq) allows simultaneous delineation of in vivo RNA secondary structure and protein-protected sites (PPSs) (Fig. 2) (Gosai et al. 2015). To identify PPSs, samples are treated with a single-strand specific or double-strand specific RNase. Proteins are then denatured before library preparation. To determine the RNA secondary structure, proteins are denatured by SDS and removed by protease digestion to make sites protected by proteins in vivo accessible for RNases. Collectively, motifs that are enriched in the samples used to determine protein protected sites compared to the samples used for structure determination are in vivo target sites of RBDs.

Gregory and coworkers applied PIP-seq to the nuclei of two specific cell types in the Arabidopsis roots that derive from epidermal cells through distinct differentiation, those cells bearing root hairs and those that do not (Foley et al. 2017). Distinct protein binding patterns were detected, and binding motifs either specific to hair cells, non-hair cells or common to both cell types were determined. To identify candidate proteins, RNA affinity chromatography was performed on immobilized oligonucleotides derived from enriched motifs. A GGN repeat motif enriched in sites protected in both hair cells and non-hair cells recovered SERRATE (SE) from root lysates, a zinc finger containing RBP involved in processing of miRNA precursors. A TG rich motif enriched in hair cell-specific protected sites identified AtGRP2, *At*GRP7 and *At*GRP8. Subsequently, *At*GRP8 was shown to regulate root hair development at the posttranscriptional level.
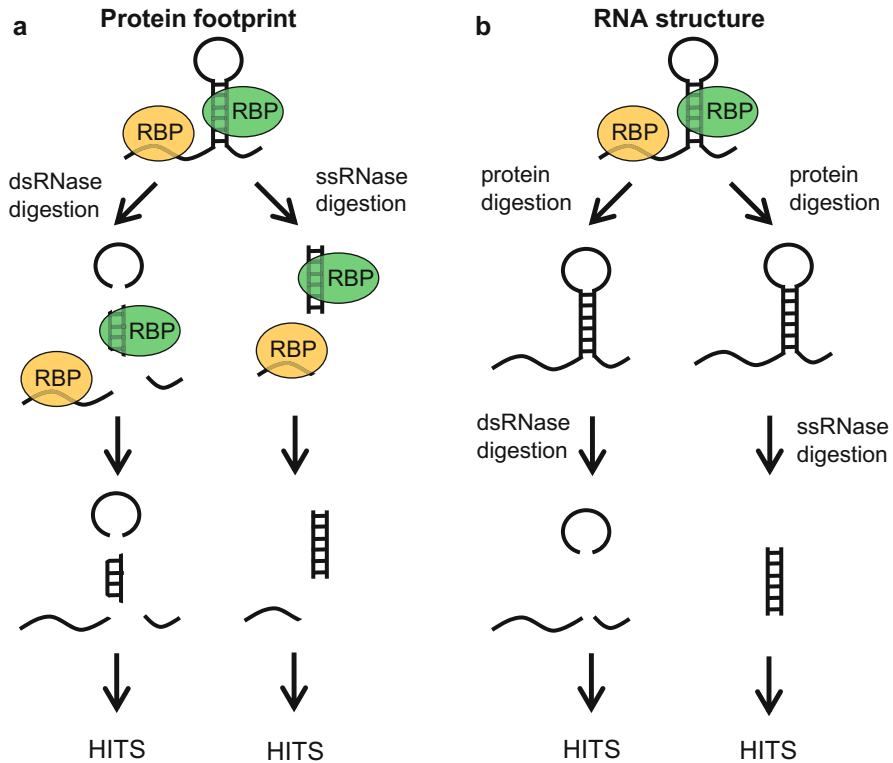
Fig. 2 Protein interaction profile sequencing (PIP-seq). (**a**) To identify protein binding sites, i.e., sites that are protected from RNase digestion by interacting proteins (PPS), samples are treated with an RNase specific for double-stranded RNA (left) or for single-stranded RNA (right). Subsequently, proteins are denatured, leaving either target sites for proteins with a preference for single-stranded regions (left), or target sites for proteins with a preference for double-stranded regions (right). These sequences are used to generate libraries for HITS. (**b**) To determine the RNA secondary structure, proteins are denatured in a first step. Subsequently, samples are treated with RNase specific for double-stranded RNA (left) or for single-stranded RNA (right). Again, libraries for HITS are prepared. Collectively, motifs that are enriched in the samples used to determine protein binding sites compared to the samples used for structure determination are in vivo target sites of RBDs

An advantage of PIP-seq is that it does not rely on an antibody to identify target sites within bound transcripts. In contrast, subsequent identification of the cognate binding proteins requires in vitro binding techniques. Thus, binding in vivo has to be confirmed by independent means.

5 Achievements and Limitations of Arabidopsis In Vivo RNA–Protein Interaction

The recent mRNA interactome capture studies are very valuable in having established UV cross-linking and oligo(dT) affinity capture to determine the mRNA binding proteome also in Arabidopsis. A large number of previously predicted RBPs in Arabidopsis were now identified experimentally and many novel proteins without a previous assignment to RNA biology unearthed. Reichel and colleagues noticed a bias toward proteins with higher abundance in the interactome compared to the input (Reichel et al. 2016), suggesting that additional proteins with lower expression level may still be identified in the future. Only few of the mRNA interacting proteins were present in all three interactomes (Köster et al. 2017). This may partly be attributed to the widely differing developmental stages investigated. Among the commonly identified proteins are numerous cytoplasmic ribosomal proteins from the small and large ribosomal subunits, likely due to their high abundance, as well as the ubiquitously expressed glycine-rich RBPs *At*GRP7 and *At*GRP8 (Köster et al. 2017).

Future applications are the dynamics of posttranscriptional networks in response to endogenous and exogenous stimuli cues by describing changes in the mRNA bound proteomes. Furthermore, as proteins binding to nonpolyadenylated RNAs obviously remain elusive in these approaches, transcript-specific approaches have to be developed.

Transcriptome-wide identification of target transcripts bound by selected RBPs in vivo has overcome a major limitation in research on plant RNA-based regulation. Nevertheless, except for the PPR proteins, we are still far from understanding the exact binding specificity of most proteins and the consequences in vivo binding has for the targets. To correlate in vivo binding with function, the impact of mutated candidate binding motifs on RBP binding and target gene expression has to be determined.

Most bioinformatics pipelines today discussing motif discovery are limited to sequence data. Current efforts focus on developing bioinformatics pipelines for identifying conserved motifs taking RNA structure context into consideration (Maticzka et al. 2014). Molecular dynamics of RNA molecules are still compute intensive but can shed light on possible interaction sites and three dimensional structures (Tuszynska et al. 2015; Boniecki et al. 2016). Finally, heterogeneous datasets and analyses, fusing several kinds of sources, can improve meta-analysis with in silico and in vivo datasets. This is yet limited in Arabidopsis but will improve the information quality in the near future. Additionally, it will be important to have comprehensive databases on RBP target sites linked to the Arabidopsis information portal (The International Arabidopsis Informatics Consortium 2012). Such resources will be of great value to improve a systems understanding of RNA–protein interaction.

References

Ali GS, Palusa SG, Golovkin M et al (2007) Regulation of plant developmental processes by a novel splicing factor. PLoS One 2:e471

Badolato J, Gardiner E, Morrison N et al (1995) Identification and characterisation of a novel human RNA-binding protein. Gene 166:323–327

Baltz AG, Munschauer M, Schwanhäusser B et al (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. Mol Cell 46:674–690

Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol 65:415–442

Beckmann BM, Horos R, Fischer B et al (2015) The RNA-binding proteomes from yeast to man harbour conserved enigmRBPs. Nat Commun 6:10127

Boniecki MJ, Lach G, Dawson WK et al (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. Nucleic Acids Res 44:e63–e63

Bunnik EM, Batugedara G, Saraf A et al (2016) The mRNA-bound proteome of the human malaria parasite Plasmodium falciparum. Genome Biol 17:147

Carvalho RF, Carvalho SD, Duque P (2010) The plant-specific SR45 protein negatively regulates glucose and ABA signaling during early seedling development in Arabidopsis. Plant Physiol 154:772–783

Carvalho RF, Szakonyi D, Simpson CG et al (2016) The Arabidopsis SR45 splicing factor, a negative regulator of sugar signaling, modulates SNF1-related protein kinase 1 stability. Plant Cell 28:1910–1925

Castello A, Fischer B, Eichelbaum K et al (2012) Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell 149:1393–1406

Cruz JA, Westhof E (2009) The dynamic landscapes of RNA architecture. Cell 136:604–609

Dong Z, Han MH, Fedoroff N (2008) The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc Natl Acad Sci USA 105:9970–9975

Ferrari R, Tadini L, Moratti F et al (2017) CRP1 Protein: (dis)similarities between *Arabidopsis thaliana* and Zea mays. Front Plant Sci 8:163

Foley SW, Gosai SJ, Wang D et al (2017) A global view of RNA-protein interactions identifies post-transcriptional regulators of root hair cell fate. Dev Cell 41:204–220

Fu ZQ, Guo M, Jeong BR et al (2007) A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. Nature 447:284–288

Galgano A, Gerber AP (2011) RNA-binding protein immunopurification-microarray (RIP-Chip) analysis to profile localized RNAs. Methods Mol Biol 714:369–385

Golovkin M, Reddy AS (1999) An SC35-like protein and a novel serine/arginine-rich protein interact with *Arabidopsis* U1-70K protein. J Biol Chem 274:36428–36438

Gosai S, Foley Shawn W, Wang D et al (2015) Global analysis of the RNA-protein interaction and RNA secondary structure landscapes of the *Arabidopsis* nucleus. Mol Cell 57:829–845

Goyal M, Banerjee C, Nag S et al (2016) The Alba protein family: structure and function. Biochim Biophys Acta Proteins Proteomics 1864:570–583

Guerreiro A, Deligianni E, Santos J et al (2014) Genome-wide RIP-chip analysis of translational repressor-bound mRNAs in the Plasmodium gametocyte. Genome Biol 15:493

Hackmann C, Korneli C, Kutyniok M et al (2014) Salicylic acid-dependent and -independent impact of an RNA-binding protein on plant immunity. Plant Cell Environ 37:696–706

Heintzen C, Melzer S, Fischer R et al (1994) A light- and temperature-entrained circadian clock controls expression of transcripts encoding nuclear proteins with homology to RNA-binding proteins in meristematic tissue. Plant J 5:799–813

Hornyik C, Terzi LC, Simpson GG (2010) The spen family protein FPA controls alternative cleavage and polyadenylation of RNA. Dev Cell 18:203–213

Hugouvieux V, Kwak JM, Schroeder JI (2001) An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in *Arabidopsis*. Cell 106:477–487

Jeong B, Lin Y, Joe A et al (2011) Structure function analysis of an ADP-ribosyltransferase type III effector and its RNA-binding target in plant immunity. J Biol Chem 286:43272–43281

Jones MA, Williams BA, McNicol J et al (2012) Mutation of *Arabidopsis* SPLICEOSOMAL TIMEKEEPER LOCUS1 causes circadian clock defects. Plant Cell 24:4907–4916

Juntawong P, Sorenson R, Bailey-Serres J (2013) Cold shock protein 1 chaperones mRNAs during translation in *Arabidopsis thaliana*. Plant J 74:1016–1028

Kalyna M, Lopato S, Barta A (2003) Ectopic expression of at RSZ33 reveals its function in splicing and causes pleiotropic changes in development. Mol Biol Cell 14:3565–3577

Keene JD (2007) RNA regulons: coordination of post-transcriptional events. Nat Rev Genet 8:533–543

Kim JS, Park SJ, Kwak KJ et al (2007a) Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in Escherichia coli. Nucleic Acids Res 35:506–516

Kim JY, Park SJ, Jang B et al (2007b) Functional characterization of a glycine-rich RNA-binding protein 2 in *Arabidopsis thaliana* under abiotic stress conditions. Plant J 50:439–451

Kim YO, Pan S, Jung CH et al (2007c) A zinc finger-containing glycine-rich RNA-binding protein, at RZ-1a, has a negative impact on seed germination and seedling growth of *Arabidopsis thaliana* under salt or drought stress conditions. Plant Cell Physiol 48:1170–1181

Kim JS, Kim KA, Oh TR et al (2008) Functional characterization of DEAD-box RNA helicases in *Arabidopsis thaliana* under abiotic stress conditions. Plant Cell Physiol 49:1563–1571

Kim JY, Kim WY, Kwak KJ et al (2010) Glycine-rich RNA-binding proteins are functionally conserved in *Arabidopsis thaliana* and Oryza sativa during cold adaptation process. J Exp Bot 61:2317–2325

König J, Zarnack K, Rot G et al (2010) iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution. Nat Struct Mol Biol 17:909–915

Köster T, Staiger D (2014) RNA-binding protein Immunoprecipitation from whole-cell extracts. Methods Mol Biol 1062:679–695

Köster T, Meyer K, Weinholdt C et al (2014) Regulation of pri-miRNA processing by the hnRNP-like protein *At*GRP7 in Arabidopsis. Nucleic Acids Res 42:9925–9936

Köster T, Marondedze C, Meyer K et al (2017) RNA-binding proteins revisited: the emerging Arabidopsis mRNA interactome. Trends Plant Sci 22:512–526

Krause K, Herrmann U, Fuß J et al (2009) Whirly proteins as communicators between plant organelles and the nucleus? Endocytobiosis Cell Res 19:51–62

Kupsch C, Ruwe H, Gusewski S et al (2012) *Arabidopsis* chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. Plant Cell 24:4266–4280

Kwon SC, Yi H, Eichelbaum K et al (2013) The RNA-binding protein repertoire of embryonic stem cells. Nat Struct Mol Biol 20:1122–1130

Lewinski M, Hallmann A, Staiger D (2016) Genome-wide identification and phylogenetic analysis of plant RNA binding proteins comprising both RNA recognition motifs and contiguous glycine residues. Mol Gen Genomics 291:763–773

Li D, Zhang H, Hong Y et al (2014) Genome-wide identification, biochemical characterization, and expression analyses of the YTH domain-containing RNA-binding protein family in *Arabidopsis* and Rice. Plant Mol Biol Report 32:1169–1186

Liao Y, Castello A, Fischer B et al (2016) The cardiomyocyte RNA-binding proteome: links to intermediary metabolism and heart disease. Cell Rep 16:1456–1469

Licatalosi DD, Mele A, Fak JJ et al (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456:464–469

Löhr B, Streitner C, Steffen A et al (2014) A glycine-rich RNA-binding protein affects gibberellin biosynthesis in *Arabidopsis*. Mol Biol Rep 41:439–445

Lopato S, Kalyna M, Dorner S et al (1999) atSRp30, one of two SF2/ASF-like proteins from *Arabidopsis thaliana*, regulates splicing of specific plant genes. Genes Dev 13:987–1001

Lorkovic ZJ, Wieczorek Kirk DA, Klahre U et al (2000) RBP45 and RBP47, two oligouridylate-specific hnRNP-like proteins interacting with poly(A)+ RNA in nuclei of plant cells. RNA 6:1610–1624

Lu C, Fedoroff N (2000) A mutation in the Arabidopsis HYL1 gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell 12:2351–2366

Lueong S, Merce C, Fischer B et al (2016) Gene expression regulatory networks in *Trypanosoma brucei*: insights into the role of the mRNA-binding proteome. Mol Microbiol 100:457–471

Lyons R, Iwase A, Gänsewig T et al (2013) The RNA-binding protein FPA regulates flg22-triggered defense responses and transcription factor activity by alternative polyadenylation. Sci Rep 3:2866

Macknight R, Bancroft I, Page T et al (1997) FCA, a gene controlling flowering time in Arabidopsis, encodes a protein containing RNA-binding domains. Cell 89:737–745

Marondedze C, Thomas L, Serrano NL et al (2016) The RNA-binding protein repertoire of *Arabidopsis thaliana*. Sci Rep 6:29766

Matia-Gonzalez AM, Laing EE, Gerber AP (2015) Conserved mRNA-binding proteomes in eukaryotic organisms. Nat Struct Mol Biol 22:1027–1033

Maticzka D, Lange S, Costa F et al (2014) GraphProt: modeling binding preferences of RNA-binding proteins. Genome Biol 15:R17

Meyer K, Köster T, Nolte C et al (2017) Adaptation of iCLIP to plants determines the binding landscape of the clock-regulated RNA-binding protein AtGRP7. Genome Biol 18:204

Mitchell SF, Jain S, She M et al (2013) Global analysis of yeast mRNPs. Nat Struct Mol Biol 20:127–133

Müller-McNicoll M, Botti V, de Jesus Domingues AM et al (2016) SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. Genes Dev 30:553–566

Nandan D, Thomas SA, Nguyen A et al (2017) Comprehensive identification of mRNA-binding proteins of *Leishmania donovani* by interactome capture. PLoS One 12:e0170068

Nicaise V, Joe A, Jeong B et al (2013) Pseudomonas HopU1 affects interaction of plant immune receptor mRNAs to the RNA-binding protein GRP7. EMBO J 32:701–712

Ohta M, Sugita M, Sugiura M (1995) Three types of nuclear genes encoding chloroplast RNA-binding proteins (cp29, cp31 and cp33) are present in *Arabidopsis thaliana*: presence of cp31 in chloroplasts and its homologue in nuclei/cytoplasms. Plant Mol Biol 27:529–539

Park SJ, Kwak KJ, Oh TR et al (2009) Cold shock domain proteins affect seed germination and growth of *Arabidopsis thaliana* under abiotic stress conditions. Plant Cell Physiol 50:869–878

Perez-Santángelo S, Mancini E, Francey LJ et al (2014) Role for LSM genes in the regulation of circadian rhythms. Proc Natl Acad Sci USA 111:15166–15171

Prikryl J, Watkins KP, Friso G et al (2008) A member of the Whirly family is a multifunctional RNA- and DNA-binding protein that is essential for chloroplast biogenesis. Nucleic Acids Res 36:5152–5165

Qi Y, Tsuda K, Joe A et al (2010) A putative RNA-binding protein positively regulates salicylic acid-mediated immunity in *Arabidopsis*. Mol Plant Microbe Interact 23:1573–1583

Reichel M, Liao Y, Rettel M et al (2016) In planta determination of the mRNA-binding proteome of Arabidopsis etiolated seedlings. Plant Cell 28:2435–2452

Ren G, Xie M, Dou Y et al (2012) Regulation of miRNA abundance by RNA binding protein TOUGH in *Arabidopsis*. Proc Natl Acad Sci USA 109:12817–12821

Riera M, Redko Y, Leung J (2006) *Arabidopsis* RNA-binding protein UBA2a relocalizes into nuclear speckles in response to abscisic acid. FEBS Lett 580:4160–4165

Ripoll JJ, Ferrandiz C, Martinez-Laborda A et al (2006) PEPPER, a novel K-homology domain gene, regulates vegetative and gynoecium development in *Arabidopsis*. Dev Biol 289:346–359

Rossbach O, Hung L-H, Khrameeva E et al (2014) Crosslinking-immunoprecipition (iCLIP) analysis reveals global regulatory roles of hnRNP L. RNA Biol 11:146–155

Rudolf F, Wehrle F, Staiger D (2004) Slave to the rhythm. Biochemist 26:11–13

Rühl C, Stauffer E, Kahles A et al (2012) Polypyrimidine tract binding protein homologs from *Arabidopsis* are key regulators of alternative splicing with implications in fundamental developmental processes. Plant Cell 24:4360–4375

Ruwe H, Kupsch C, Teubner M et al (2011) The RNA-recognition motif in chloroplasts. J Plant Physiol 168:1361–1371

Schmal C, Reimann P, Staiger D (2013) A circadian clock-regulated toggle switch explains AtGRP7 and AtGRP8 oscillations in *Arabidopsis thaliana*. PLoS Comput Biol 9:e1002986

Schmidt F, Marnef A, Cheung M-K et al (2010) A proteomic analysis of oligo(dT)-bound mRNP containing oxidative stress-induced *Arabidopsis thaliana* RNA-binding proteins *AT*GRP7 and *AT*GRP8. Mol Biol Rep 37:839–845

Schmitz-Linneweber C, Small I (2008) Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci 13:663–670

Silverman IM, Li F, Gregory BD (2013) Genomic era analyses of RNA secondary structure and RNA-binding proteins reveal their significance to post-transcriptional regulation in plants. Plant Sci 205-206:55–62

Simpson GG, Dijkwel PP, Quesada V et al (2003) FY is an RNA 3′ end-processing factor that interacts with FCA to control the *Arabidopsis* floral transition. Cell 113:777–787

Sommerville J (1999) Activities of cold-shock domain proteins in translation control. BioEssays 21:319–325

Sorenson R, Bailey-Serres J (2014) Selective mRNA sequestration by OLIGOURIDYLATE-BINDING PROTEIN 1 contributes to translational control during hypoxia in *Arabidopsis*. Proc Natl Acad Sci USA 111:2373–2378

Staiger D (2001) RNA-binding proteins and circadian rhythms in *Arabidopsis thaliana*. Philos Trans R Soc Lond Ser B Biol Sci 356:1755–1759

Staiger D (2015) Shaping the Arabidopsis transcriptome through alternative splicing. Adv Bot 2015:13

Staiger D, Zecca L, Wieczorek Kirk DA et al (2003) The circadian clock regulated RNA-binding protein AtGRP7 autoregulates its expression by influencing alternative splicing of its own pre-mRNA. Plant J 33:361–371

Stauffer E, Westermann A, Wagner G et al (2010) Polypyrimidine tract-binding protein homologues from Arabidopsis underlie regulatory circuits based on alternative splicing and downstream control. Plant J 64:243–255

Streitner C, Danisman S, Wehrle F et al (2008) The small glycine-rich RNA-binding protein *At*GRP7 promotes floral transition in *Arabidopsis thaliana*. Plant J 56:239–250

Streitner C, Köster T, Simpson CG et al (2012) An hnRNP-like RNA-binding protein affects alternative splicing by in vivo interaction with target transcripts in *Arabidopsis thaliana*. Nucleic Acids Res 40:11240–11255

Sysoev VO, Fischer B, Frese CK et al (2016) Global changes of the RNA-bound proteome during the maternal-to-zygotic transition in Drosophila. Nat Commun 7:12128

Tenenbaum SA, Carson CC, Lager PJ et al (2000) Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. Proc Natl Acad Sci USA 97:14085–14090

Teubner M, Fuß J, Kühn K et al (2017) The RRM protein CP33A is a global ligand of chloroplast mRNAs and is essential for plastid biogenesis and plant development. Plant J 89:472–485

The International Arabidopsis Informatics Consortium (2012) Taking the next step: building an Arabidopsis information portal. Plant Cell 24:2248–2256

Tillich M, Hardel SL, Kupsch C et al (2009) Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. Proc Natl Acad Sci USA 106:6002–6007

Tuszynska I, Magnus M, Jonak K et al (2015) NPDock: a web server for protein–nucleic acid docking. Nucleic Acids Res 43:W425–W430

Vandivier LE, Anderson SJ, Foley SW et al (2016) The conservation and function of RNA secondary structure in plants. Annu Rev Plant Biol 67:463–488

Vazquez F, Gasciolli V, Crete P et al (2004) The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol 14:346–351

Völz R, von Lyncker L, Baumann N et al (2012) LACHESIS-dependent egg-cell signaling regulates the development of female gametophytic cells. Development 139:498–502

Wang X, Lu Z, Gomez A et al (2014) N^6-methyladenosine-dependent regulation of messenger RNA stability. Nature 505:117–120

Wessels H-H, Imami K, Baltz AG et al (2016) The mRNA-bound proteome of the early fly embryo. Genome Res 26:1000–1009

Xing D, Wang Y, Hamilton M et al (2015) Transcriptome-wide identification of RNA targets of Arabidopsis SERINE/ARGININE-RICH45 uncovers the unexpected roles of this RNA binding protein in RNA processing. Plant Cell 27:3294–3308

Zhang X-N, Mount SM (2009) Two alternatively spliced isoforms of the *Arabidopsis thaliana* SR45 protein have distinct roles during normal plant development. Plant Physiol 150:1450–1458

Zhang Y, Gu L, Hou Y et al (2015) Integrative genome-wide analysis reveals HLP1, a novel RNA-binding protein, regulates plant flowering by targeting alternative polyadenylation. Cell Res 25:864–876

Zhang Z, Boonen K, Ferrari P et al (2016) UV crosslinked mRNA-binding proteins captured from leaf mesophyll protoplasts. Plant Methods 12:42

Zhang X-N, Shi Y, Powers JJ et al (2017) Transcriptome analyses reveal SR45 to be a neutral splicing regulator and a suppressor of innate immunity in *Arabidopsis thaliana*. BMC Genomics 18:772